geneGIS: Computational Tools for Spatial Analyses of DNA Profiles with Associated Photo-Identification and Telemetry Records of Marine Mammals

C. Scott Baker
Oregon State University
Hatfield Marine Science Center
Newport, OR 97365-5296
phone: (541) 867-0255 fax: (541) 867-0138 email: scott.baker@oregonstate.edu

Dawn Wright
Department of Geosciences
Oregon State University
phone: (541) 737-1229 fax: (541) 737-1200 email: dawn@science.oregonstate.edu

John Calambokidis
Cascadia Research
218 ½ W. 4th Ave.
Olympia, WA 98501
phone: (360) 943-7325 ext 104 email: calambokidis@cascadiaresearch.org

Award Number: N000141110614
http://mmi.oregonstate.edu/ccgl

LONG-TERM GOALS

We are developing computational tools for improved visual exploration and spatial analysis of DNA profiles, with accompanying photo-identification records or telemetry tracks of marine mammals. Referred to as *geneGIS*, the computational tools provide the ability to display, browse, select, filter and summarize spatio-temporal relationships of these individual-based records and associated data from molecular markers and ecomarkers (e.g., stable isotopes). A toolbox of software applications allows basic summaries of spatially selected data and export of data in standard tabular and database formats (e.g., XLS, CSV, MDB, KML), as well as specialized formats required for programs commonly used in molecular ecology and capture-mark-recapture. The data format complies with OBIS standards and the database architecture is compatible with the Arc Marine data model, providing a link with other datasets and tools needed for an integrated description of the genetic and environmental 'seascape' of cetaceans. We have implemented *geneGIS* as toolboxes in the desktop version of ArcGIS 10.1 and through programmatic enhancements of the web-based *Shepherd Project* using DNA profiles and photo-identification records derived from an ocean-wide survey of humpback whales in the North Pacific (*SPLASH* and *geneSPLASH*). We have recently been granted access to a subset of photo-identification sighting records and associated DNA profiles from the North Atlantic Right Whale Consortium for implementation in *geneGIS*. Spatio-temporal analyses of databases from long-lived, migratory whales will be suitable for informing conceptual models of cetacean populations, including

Report Documentation Page			*Form Approved* *OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.			
1. REPORT DATE **2012**	2. REPORT TYPE **N/A**		3. DATES COVERED **-**
4. TITLE AND SUBTITLE **geneGIS: Computational Tools for Spatial Analyses of DNA Profiles with Associated Photo-Identification and Telemetry Records of Marine Mammals**			5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)			5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Oregon State University Hatfield Marine Science Center Newport, OR 97365-5296**			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited**			
13. SUPPLEMENTARY NOTES **The original document contains color images.**			
14. ABSTRACT			
15. SUBJECT TERMS			

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**SAR**	**16**	

the Population Consequences of Acoustic Disturbance (PCAD) and the Testing of Spatial Structure Methods (TOSSM).

OBJECTIVES

The overall objectives can be ordered into five tasks (with related subtasks):

- Task 1: Develop database architecture following Arc Marine data model for integration and display of DNA profiles with photo-identification and telemetry records in a stand-alone ArcGIS framework, and enhanced features of a web-based application (*Shepherd Project*) currently designed for display and visual exploration of photo-identification catalogues.
- Task 2: Develop ArcGIS tools for data query, visual exploration and basic statistical summaries for spatial and temporal partitions of individual-based records (DNA profiles, photo-identification records and telemetry tracks).
- Task 3: Enhance user-directed spatial/temporal selection and export of individual-based records for advanced statistical analyses. This will include tools to export data compatible with existing software used for genetic analyses and capture-mark-recapture.
- Task 4: Demonstrate functionality of *geneGIS* and web-based application through importation and integration of existing large-scale datasets of photo-identification records from the Structure of Populations, Levels of Abundance and Status of Humpbacks program in the North Pacific (*SPLASH*) and associated DNA profiles from biopsy samples (*geneSPLASH*).
- Task 5: Prepare a comprehensive user guide for all software functions and analyses implemented in the system.

APPROACH

The computational developments of *geneGIS* follow two approaches: 1) tools within a web-based program for displaying individual identification photographs and information from linked DNA profiles, including a graphical user interface (GUI) through a Google Map interface; and 2) tools within ArcGIS, the most widely available software for GIS and advanced spatial analysis. The intent is to benefit from the strengths of each approach while assuring compatibiliy and interoperability through a common database architecture and simplified import/export functions.

The web-based approach is being developed under subcontract to Jason Holmberg of the *Shepherd Project*, with support of John Calambokidis and Erin Falcone of Cascadia Research Collective. This approach takes advantage of an existing open-source software framework supporting capture-mark-recapture (CMR) studies of marine megafauna by the *Shepherd Project* (Holmberg *et al*. 2008). This software framework provides a scalable, web-based platform for CMR data management and was selected by Cascadia Research Collective to develop and host the SPLASH photo-identification catalog (available in beta version as http://www.splashcatalog.org). With support from the *geneGIS* initiative, the software framework of the *Shepherd Project* has been modified to include DNA profiles, providing a computational environment suitable for integrated studies of molecular ecology and CMR (http://www.splashcatalog.org/latestgenegis/).

The ArcGIS approach is being directed by the PI through Oregon State University(OSU), with support from Professor Dawn Wright (on leave from OSU as Esri Chief Scientist), her PhD student at OSU, Dori Dick, her research assistant at Esri, Shaun Walbridge, and members of the Marine Mammal Institute (MMI), including Tomas Follet and Debbie Steel. This approach takes advantage of previous experience with management of a whale telemetry database under the Arc Marine data model (Lord-Castillo *et al.* 2009; Wright *et al.* 2007). With support from the *geneGIS* initiative, we are developing tools to import and visualize spatial distributions and selection of individual identification records, as well as raster-based data extraction from environmental layers available in the ArcGIS environment.

WORK COMPLETED

1) Integrated database of photo-identification records and DNA profiles representing the *SPLASH* and *geneSPLASH* project;
2) Import/Export functions for individual-based records into ArcGIS (SRGD.csv and Arc Marine) and *Shepherd Project*;
3) *geneGIS* tool for spatial selection and comparison in ArcGIS;
4) *geneGIS* tools for data extraction from environmental data layers in ArcGIS;
5) Implementation of integrated *SPLASH/geneSPLASH* in Java-based, web-accessible database through the *Shepherd Project*; and
6) *geneGIS* tools for custom analysis of molecular ecology and Capture-Mark-Recapture (CMR) in the *Shepherd Project*.

RESULTS

Integrated SPLASH/geneSPLASH databases

The *SPLASH* program provided a comprehensive dataset for implementation within ArcGIS and the *Shepherd Project*. At the inception of this project, *SPLASH* existed as relational database in Microsoft Access with eight primary data tables containing effort, photo-identification, and tissue sampling records for humpback whales collected during five seasons of dedicated research effort in the North Pacific. This database and the associated photographic catalogue are maintained by Cascadia Research Collective. The photo-identification dataset was reconciled prior to this grant, to include over 18,400 encounters with 7,941 unique individuals and repeated encounters with individuals could be tracked throughout the database using an identifier known as the SPLASHID number (Calambokidis *et al.* 2008).

A total of 5,675 tissue samples (mostly by biopsy darting) were also collected during SPLASH, of which about half were associated with a photo-identification encounter. From this total, 2,720 samples were selected for DNA profiling, including a sex marker, mitochondrial (mt) DNA haplotype sequencing and genotyping at 10 microsatellite loci. As expected, the 10 microsatellites were sufficiently variable to provide a second source of individual identification (Constantine *et al.* 2012; Madon *et al.* 2011). From the 2,720 samples, comparison of genotypes resolved 2,161 individuals. Referred to as *geneSPLASH*, the database of DNA profiles (i.e., a ‘DNA register’ DeSalle and Amato 2004) is maintained by the Cetacean Conservation and Genomics Laboratory, MMI, OSU.

As part of the *geneGIS* initiative, these SPLASH photo-identification records and the *geneSPLASH* DNA profiles were integrated into a master *SPLASH/geneSPLASH* database (e.g., Figures 1 and 4). By review and reconciliation of disagreements in the two sources of identification, we were able to correct a small number of errors in both the photographic and genetic datasets (Falcone and Steel, personal communication). The most common sources of discrepancies were errors in sample assignment that occurred in the field, with internal matches in the photographic catalog a. In many cases, a detailed review of data linked to the sample (photos, field notes) allowed correct reassignment to another whale in the same group. Where a sample could not be confidently reassigned to another whale, the link between the sample and the whale was removed and the sample was attributed to an unknown individual. Ultimately, all but two known inconsistencies were resolved through this review process.

After the reconciliation, 1,452 of the individuals identified by DNA profiling were also identified by an associated photo-. These identification records were tracked using the original SPLASHID number allowing information from the DNA profiles to be 'extended' across multiple photo-identification encounters. For example, the DNA profile associated with a biopsy sample and a photo-identification encounter collected in Hawaii could be extended to the photo-identification records of that individual in southeastern Alaska (e.g., see Figure 4). The remaining 709 individuals identified only by a DNA profile were assigned a new SPLASHID, increasing the total catalog size from 7,941 to 8,651 unique individuals. The fully integrated database of photo-identification records and DNA profiles is one of the largest yet assembled for living whales.

Database architecture and tabular input files

A database architecture has been agreed to accommodate relational data typical of those used in the collection of individual-based records from photo-identification and telemetry, with the associated collection of tissue samples for genetic analyses and ecomarkers. The architecture and nomenclature conform to Arc Marine and Darwin Core standards (Wieczorek *et al.* 2012; Lord-Castillo *et al.* 2009; Wright *et al*. 2007), where possible, and can accommodate the current databases developed for telemetry data and DNA profiles at MMI and SPLASH records at Cascadia Research Collective.

We have also developed import options for a simplified tabular structure (e.g., SRGD.csv or Excel) similar to that more commonly used in molecular ecology and Capture-Mark-Recapture. The Spatially Referenced Genetic Data format (or SRGD) is a comma-separated file providing for spatio-temporal records of encounters with individuals and associated DNA profiles (Table 1). Additional data fields relating to the encounter or the individual, such as group size, behavioral roles or ecomarkers, can be placed after the primary fields. The intent is to allow an easy entry into ArcGIS and *Shepherd Project* for access to the *geneGIS* tools by import of files that closely matches the existing data formats used in genotyping and photo-identification.

geneGIS in ArcGIS

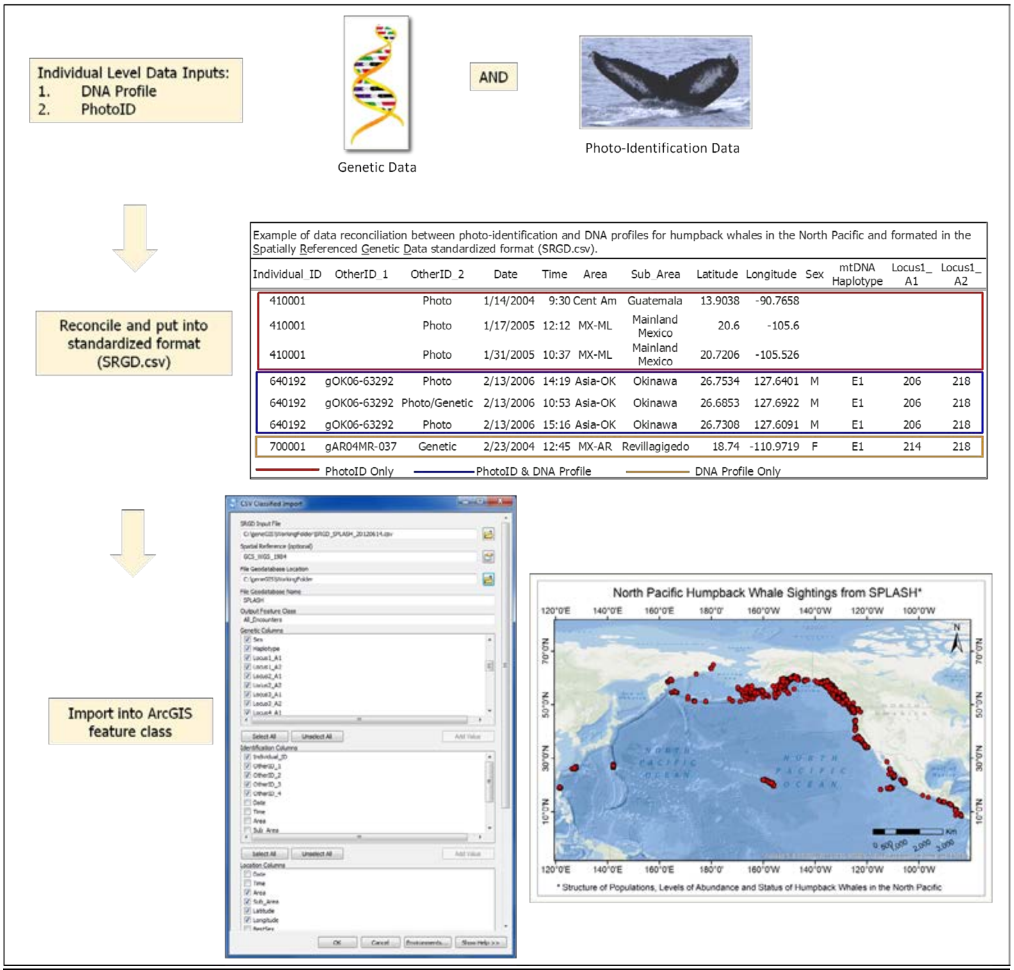
Data Visualization, Selection and Export Tools

Tool development in ArcGIS has focused on data import (see Figure 1), followed by spatial selection and export to other genetic analyses applications (e.g. GenAlEx and Genepop) (Step 3) via an ArcGIS 10.1 toolbox and a Python Addin GUI. Figure 2 illustrates a *Use Case* for the spatial selection of the reconciled *SPLASH*/*geneSPLASH* database, to compare differences in mtDNA frequencies for humpback whales sighted in the western Gulf of Alaska and in southeast Alaska, via export to GenAlEx (Peakall and Smouse 2006).

Environmental Data Extraction Tool
Once data are loaded into ArcGIS, the user may want to also add environmental layers (e.g., bathymetry, sea surface temperature) and extract values for each location where an individual was sighted. This information can be used to conduct more advanced spatial analyses and provide an improved understanding on the influence of the environment on the distribution of a species (environmental seascape of cetaceans). Figure 3 illustrates a workflow example of bathymetric data extraction for all *SPLASH* humpback whale encounters during 2004-2006.

Table 1. Required fields, descriptions and data types for the SRGD.csv input file. The fields shown are the minimum recommended data requirements to take advantage of geneGIS tools. Additional fields (e.g., group occurrence, group size, behavior role, ecomarker values) can be included following the required fields.

Field Name	Description	Data Format
Individual_ID	A unique identifier for each individual or sample in your data set	All Integer/All Text
OtherID_1	Optional additional reference number. Provide a brief description in the FieldReference Tab	All Integer/All Text
OtherID_2	Optional. Insert additional columns as needed. Provide a brief description in the FieldReference Tab	All Integer/All Text
Date	Optional. Date of sample, if known.	mm/dd/yyyy
Time	Optional. Time sample taken, if known.	hh:mm:ss
Area	Optional. General region , e.g., Mexico, Oregon	Text
Sub_Area	Optional. Specific area, e.g., Socorro, Newport	Text
Latitude	Latitude of sample in decimal degrees (e.g., 12.8006)	Double/Floating Point
Longitude	Longitude of sample in decimal degrees (e.g., -120.8570)	Double/Floating Point
Sex	Optional. Sex of the individual, if known	Text
Haplotype	Optional. Mitochondrial haplotype of the individual, if known. Provide a brief description in the FieldReference Tab	Text
Locus1_A1	Allele name 1 of Locus 1. Provide a brief description in the FieldReference tab	All Integer/All Text
Locus1_A2	Allele name 2 of Locus 1.	All Integer/All Text
Locus2_A1	Allele name 1 of Locus 2.	All Integer/All Text
Locus2_A2	Allele name 2 of Locus 2. Add as many as you need.	All Integer/All Text



Example of data reconciliation between photo-identification and DNA profiles for humpback whales in the North Pacific and formated in the Spatially Referenced Genetic Data standardized format (SRGD.csv).

Individual_ID	OtherID_1	OtherID_2	Date	Time	Area	Sub_Area	Latitude	Longitude	Sex	mtDNA Haplotype	Locus1_A1	Locus1_A2
410001		Photo	1/14/2004	9:30	Cent Am	Guatemala	13.9038	-90.7658				
410001		Photo	1/17/2005	12:12	MX-ML	Mainland Mexico	20.6	-105.6				
410001		Photo	1/31/2005	10:37	MX-ML	Mainland Mexico	20.7206	-105.526				
640192	gOK06-63292	Photo	2/13/2006	14:19	Asia-OK	Okinawa	26.7534	127.6401	M	E1	206	218
640192	gOK06-63292	Photo/Genetic	2/13/2006	10:53	Asia-OK	Okinawa	26.6853	127.6922	M	E1	206	218
640192	gOK06-63292	Photo	2/13/2006	15:16	Asia-OK	Okinawa	26.7308	127.6091	M	E1	206	218
700001	gAR04MR-037	Genetic	2/23/2004	12:45	MX-AR	Revillagigedo	18.74	-110.9719	F	E1	214	218

PhotoID Only — PhotoID & DNA Profile — DNA Profile Only

Figure 1. An example of data import to ArcGIS 10.1. Individual-based DNA profiles (geneSPLASH) and photo-identification records (SPLASH) are reconciled and individual encounter records are formatted into a SRGD.csv input file (see Table 1). The CSV Classified Import tool allows the user to dynamically classify data columns into standard fields, such as individual identification codes, spatial/temporal locations and genetic markers (Locus1, Locus2 etc), before creating a new feature class viewable in ArcGIS.

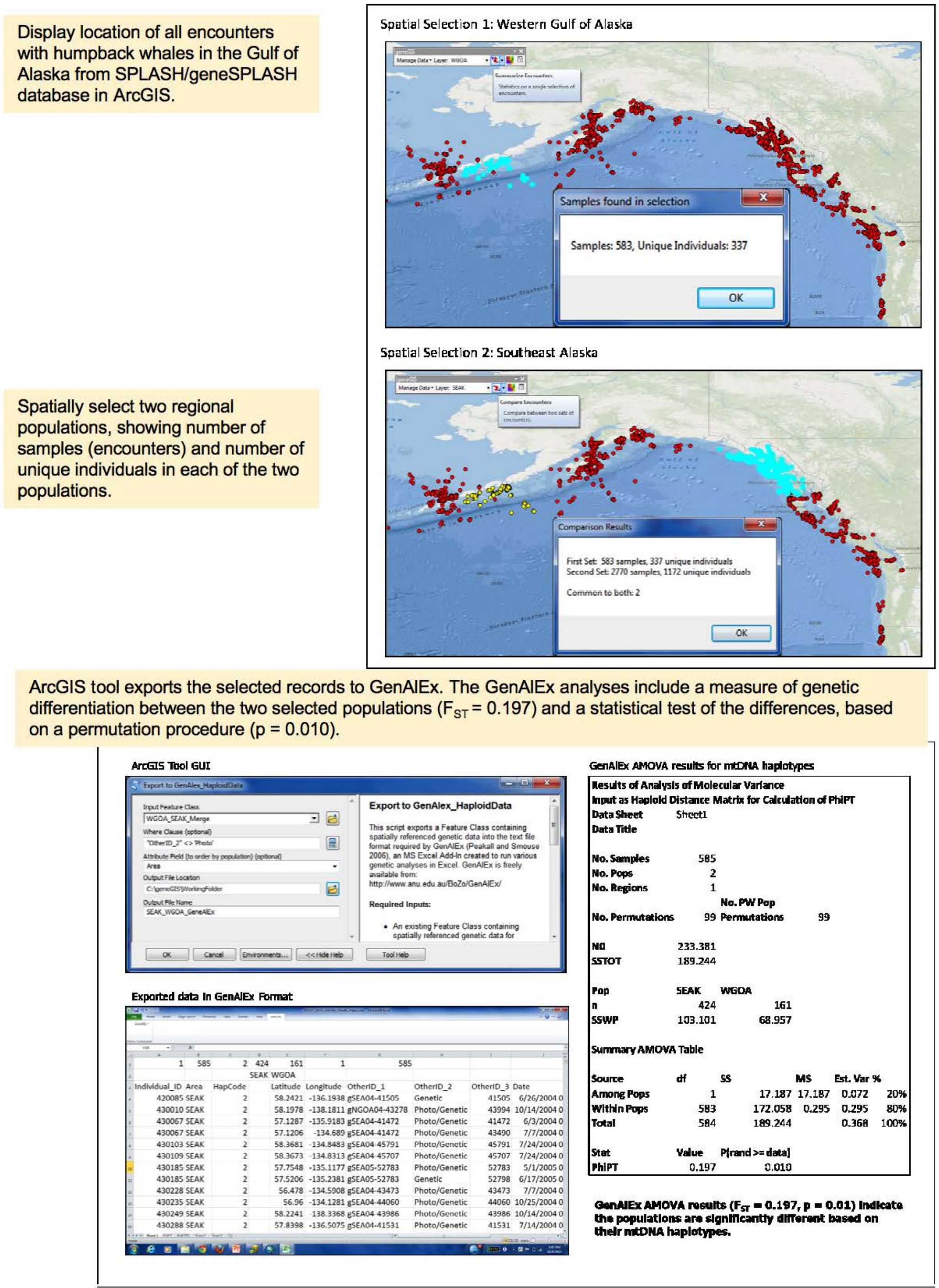


Figure 2. Use case for a spatial selection and population comparison of individually identified humpback whales encountered in the western Gulf of Alaska and southeast Alaska using geneGIS tools in ArcGIS. Spatially selected records are exported for analysis of population differentiation in the Excel Addin GenAlEx (Peakall and Smouse 2006)

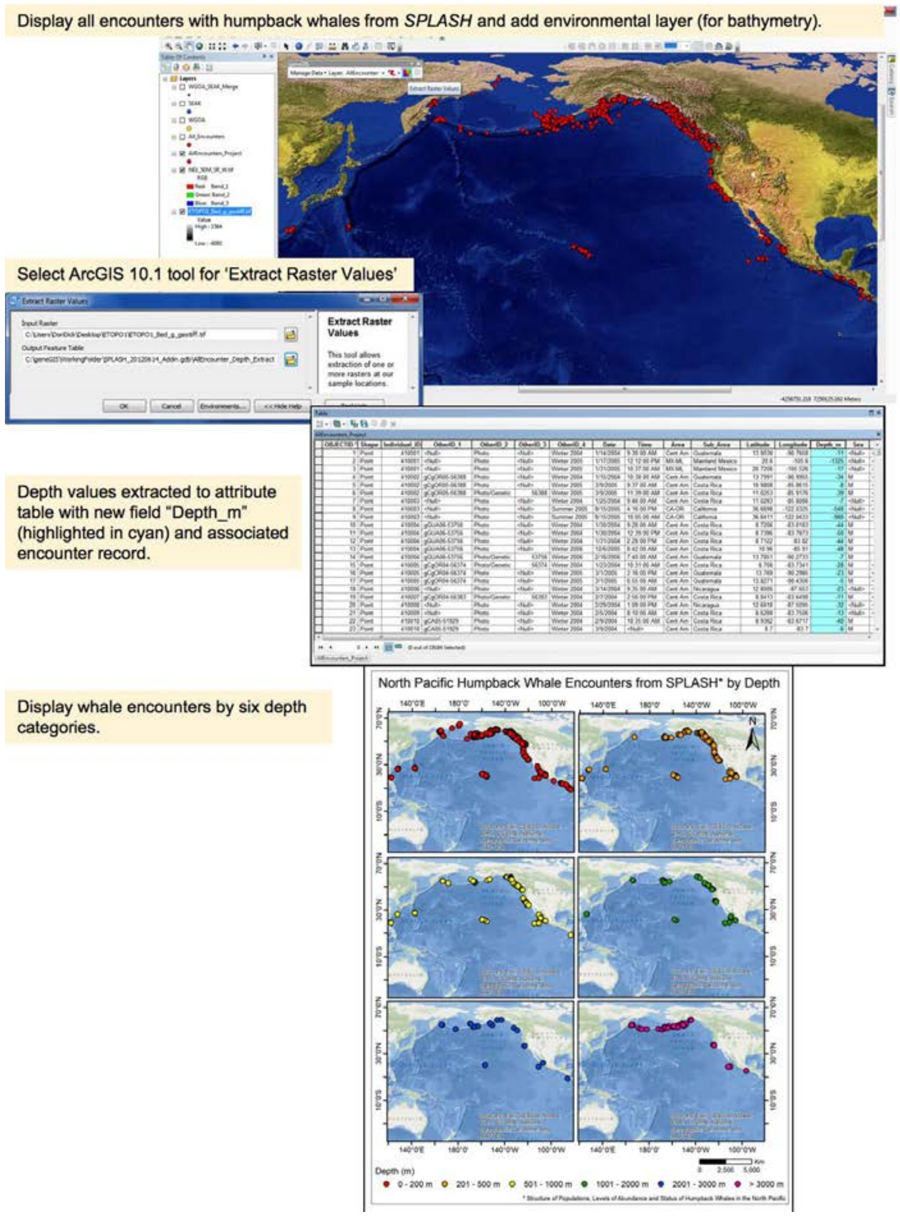


Figure 3. Use case for data extraction from an environmental layer (bathymetry) for encounter locations of humpback whales during the SPLASH project. A depth value for every humpback whale encounter during 2004-2006 is extracted from the bathymetric layer and added to each record in the point feature class. The depths for all encounters are then shown for six different depth intervals.

geneGIS in the Shepherd Project

The *Shepherd Project* (http://www.ecoceanusa.org/shepherd) is an open-source database framework designed to support studies in capture-mark-recapture (CMR), with recent enhancement to support molecular ecology and social ecology. The *Shepherd Project* is complementary to existing specialized programs for these studies, including Program MARK for estimates of population abundance and trends (White and Burnham 1999), Genepop for estimation of population genetic parameters (Rousset 2008), and SOCPROG for analysis of social affiliations (Whitehead 2009). Features of the *Shepherd Project* include:

- a scalable, collaborative platform for intelligent data storage and management, including advanced, consolidated searching;
- an easy-to-use suite of computation tools that can be extended to meet the needs of studies involving individual identification records and molecular ecology (e.g., photo-identification and DNA profiles);
- the easy export of data to specialized analysis applications (e.g., Genepop, Program Capture) and other software (e.g., Google Map); and
- the export of datasets in biodiversity databases (e.g., GBIF and OBIS);

Documentation for the complete capabilities of the *Shepherd Project* is available at:
http://www.ecoceanusa.org/shepherd

Unified Data Search and Display

The *Shepherd Project* was previously selected by Cascadia Research Collective as the 'on-line' repository for search and display of the SPLASH photo-identification catalogue. With funding from ONR, the *Shepherd Project* framework has been enhanced to support the integrated database of photo-identification and DNA profiles (Figures 4 and 5).

Spatial Selection and Custom Analysis of Population Differentiation

geneGIS tools have been implemented through six versions (2.x-3.0) to include new capabilities for supporting molecular ecology. These enhancements have resulted in improved data search capabilities and visualization of molecular markers linked to individual identification, as displayed in the following use case (Figure 6).

Marked Individual : 430045

A marked individual is a single animal distinctly identified at one or more points in time. ***Learn more.***

Tweet Like 0

Alternate ID: None [edit]

Nickname: [edit]
Nicknamed by:

Sex: male [edit]

BehSex
m1 [edit]

Best Sex Confidence
high [edit]

7 Encounter(s) (not all may be currently visible)

Date	Location	Data Types	Number	Alternate ID	Sex	Occurring With	Behavior
2006-1-19 9:06	Big Island		26574	HI06-4044	male		Singer *Group Behavior* Singing
2005-7-15 11:52	SEAK		21818	CAEP 388	male	470452 574352 460301 574002 574463 574225 474214 474279 474098 474239 474437 574395 474100 474408 *See occurrence*	Unknown
2004-9-5 15:09	Southeast Alaska		11591	1215	male	474229 470728 474350 474437 474369 700600 700601 474367 474228 *See occurrence*	Unknown *Group Behavior* Feeding
2004-6-27 14:07	Southeast Alaska		8585	1215	male		Solo *Group Behavior* Feeding
2004-6-27 15:24	Southeast Alaska		10158	1215	male	474153 *See occurrence*	Unknown *Group Behavior* Feeding
2004-6-27 15:24	Southeast Alaska		10160	1215	male	474153 *See occurrence*	Unknown *Group Behavior* Feeding
2004-2-8 9:59	Maui		17	HI04-2690	male	430085 430655 430457 430162 430349 430025 430698 430120 430232 *See occurrence*	CH/SE *Group Behavior* competition

Biological Samples

Sample ID	Encounter	Attributes	Analyses
TDI-026-04	17	Alternate Sample ID: 40293 Storage lab ID: HAWAII	**Genetic sex:** M
OT 040627-02	8585	Alternate Sample ID: 41508 Storage lab ID: AK, SEAK, CHATHAM ST., PT. AUGUSTA	**Genetic sex:** M **Haplotype:** A+ **Microsatellite Markers** GATA417: 206 214 Ev37: 204 210 Ev96: 163 163 rw4-10: 196 196 GT211: 110 110 Ev14: 131 135 rw48: 116 118 GATA28: 147 147 GT23: 115 115 GT575: 153 153
HMMC2006-016	26574	Alternate Sample ID: 52569 Storage lab ID: HAWAII, HAWAII IS.	**Genetic sex:** M

Figure 4. An example display of unified data for encounters with a marked individual using multiple sources of identification in SPLASH/geneSPLASH in the Shepherd Project. The search for individual 430045 shows a summary of 7 encounters from 2004-2006 in either southeastern Alaska or Hawaii. Photo-identification was collected from all 7 encounters. Tissue samples (biopsy or sloughed skin) were collected during encounters 26574 in Hawaii and 8585 in southeastern Alaska. The DNA profile (sex, mtDNA haplotypes and 10 microsatellite genotypes) is available from the sample collected during encounter 8585.

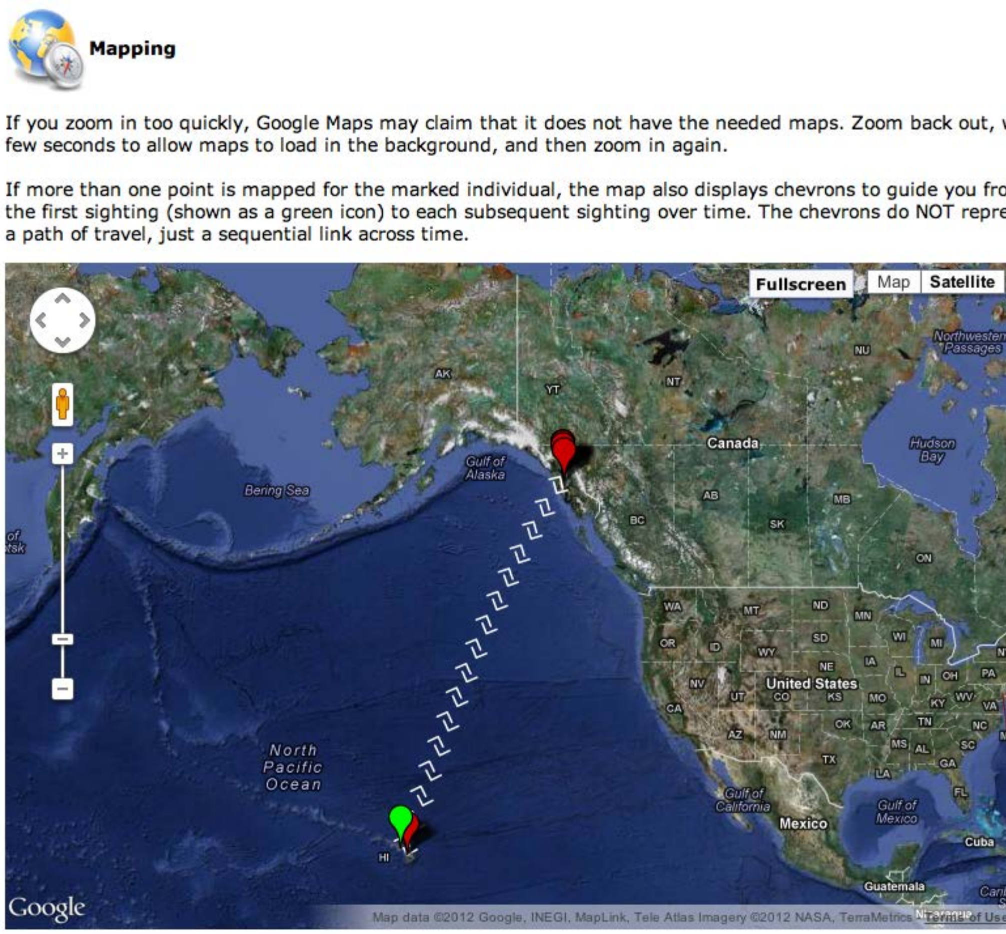


Figure 5.: An example display of unified data for encounters with marked individual using multiple sources of identification in SPLASH/geneSPLASH in Shepherd Project (continued). The Google Map shows the locations of encounters with individual 430045 from 2004-2006. The green icon shows the location of the first encounter and the arrows indicate the sequential links between encounters, showing the documented return migration between Hawaii and southeastern Alaska.

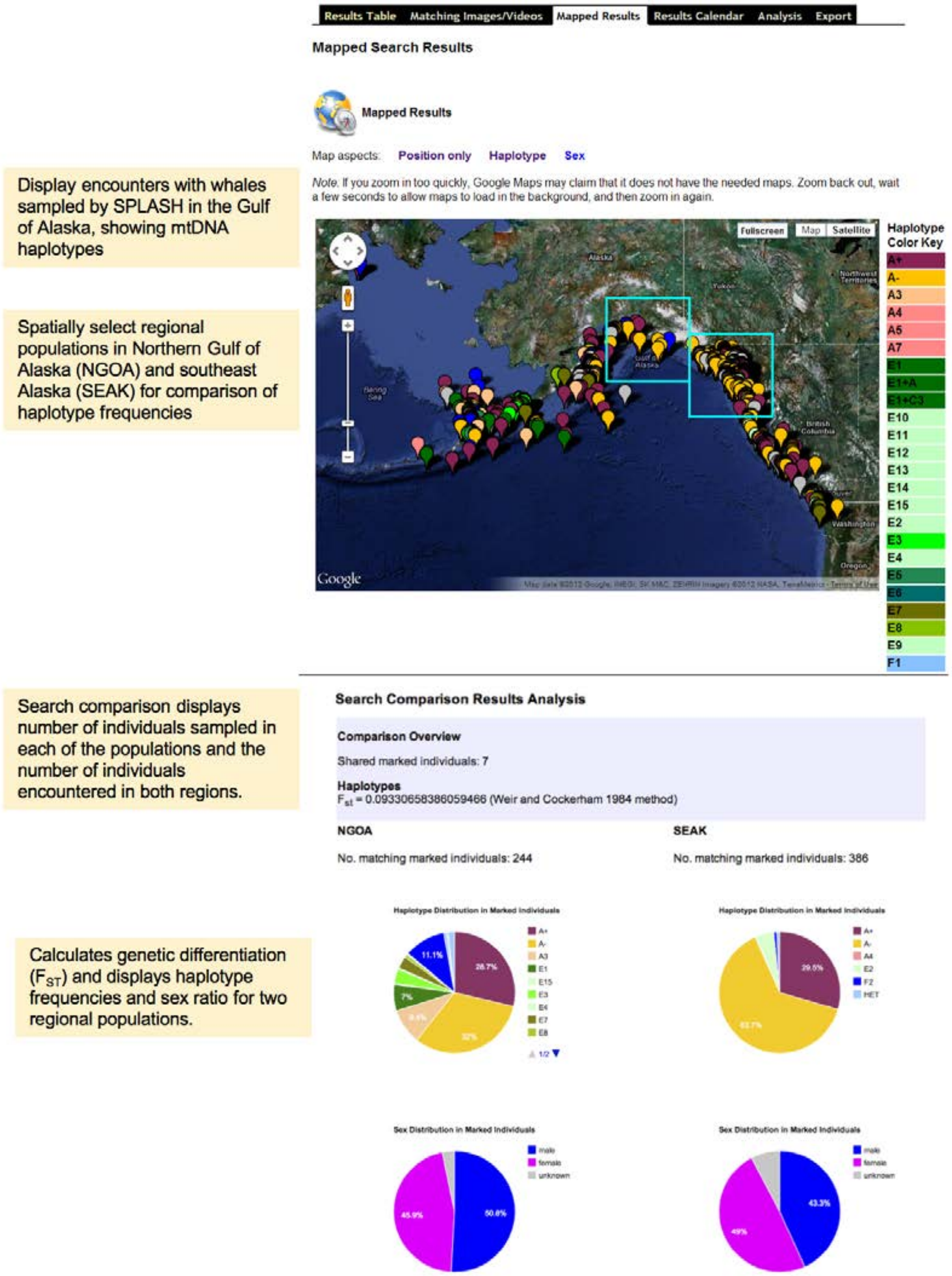


Figure 6. Use case showing encounters with individually identified humpback whales in the Gulf of Alaska during SPLASH. Each encounter is symbolized with its associated mtDNA from geneSPLASH. Spatial selection of the displayed encounters allows a customized comparison of mtDNA haplotype frequencies and sex ratios, and calculation of a standard index of genetic differentiation (F_{ST}), similar to that implemented through GenAlEx and ArcGIS.

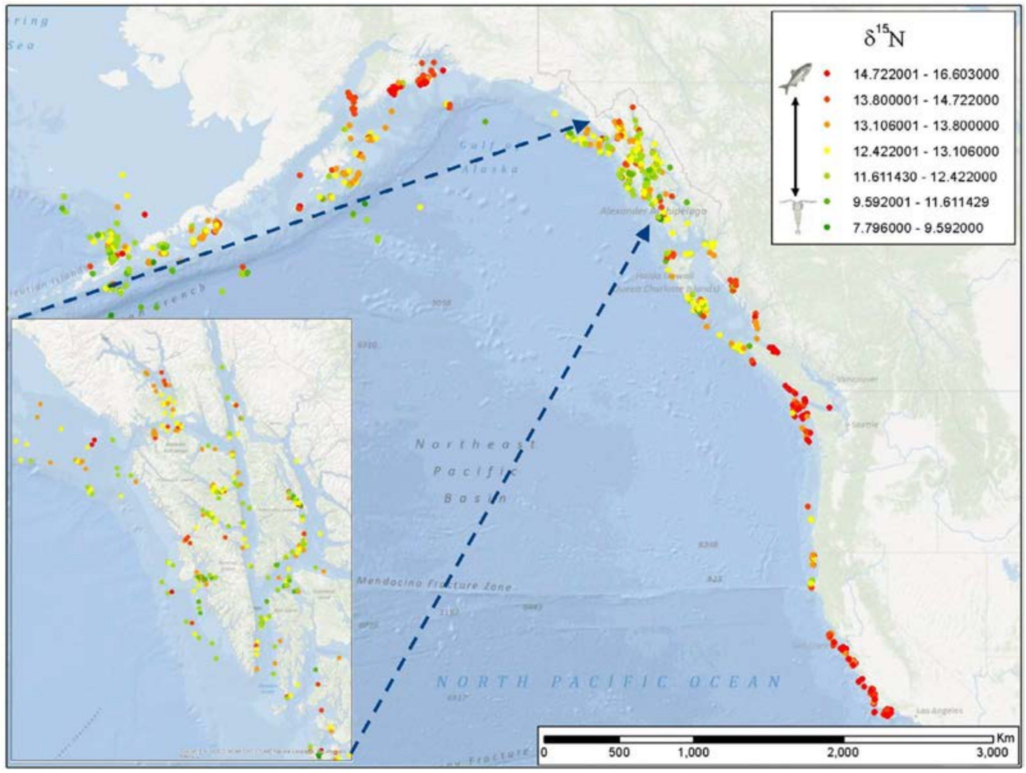


Figure 7. Geographic differences in a representative ecomarker, the stable isotope ^{15}N, from biopsy samples of humpback whales collected during the SPLASH program (Witteveen et al. 2009) and integrated into the Arc Marine data model. The scaling of values suggests differences in the primary trophic level of prey, with consumption of fish (red) dominating off Central California, the Olympic Peninsula, Washington, and inside waters of the northern Gulf of Alaska, and krill (green) dominating in southeastern Alaska. Whales in southeastern Alaska (inset) show apparent individual preferences or local area differences in consumption of fish and krill.

IMPACT/APPLICATIONS

With the development of *geneGIS* tools in ArcGIS we expect to improve access to individual-based records and associated DNA profiles to the community of spatial modelers, and to contribute to the developing fields of landscape and seascape genetics (Etherington 2011; Vandergast *et al.* 2012). With the *geneGIS* enhancements of the *Shepherd Project* we are developing a unified computational environment for Capture-Mark-Recapture and molecular ecology, including:

- distributed access to consolidated data and *geneGIS* analytic functions through a web browser, supporting distributed collaboration and the consolidation of multiple data sets;
- usable by a non-specialist, allowing non-GIS professionals to spatially explore and filter the data; and
- broader impact by contributing functionality to a open-source software framework also used by researchers for other species of marine megafauna (e.g., whale sharks and manta rays).

Now that *SPLASH/geneSPLASH* have been reconciled and implemented in both the Arc Marine model and the *Shepherd Project* framework, we are seeking additional datasets as exemplars for *geneGIS*. For this, we were recently granted access to a subset of photo-identification sighting records and DNA profiles from the North Atlantic Right Whale (NARW) database through application to the NARW Consortium. As with *SPLASH/geneSPLASH*, individual identification of NARWs has included both photo-identification and DNA profiling (Frasier *et al.* 2009), with photo-identification records maintained through the Digital Image Gathering and Information Tracking System (DIGITS) at the New England Aquarium. Communication is now ongoing with curator of the photo-identification records, Phillip Hamilton, and curator of the DNA profiles, Tim Frasier, on the details of the data loan. It is expected that this will involve several thousand sighting records and associated profiles of a few hundred individuals (Frasier *et al.* 2007).

RELATED PROJECTS

Title: '*Examination of health effects and long-term impacts of deployments of multiple tag types on blue, humpback, and gray whales in the eastern North Pacific*' with funding to Cascadia Research Collective, from the National Oceanographic Partnership Program (NOPP) and Interagency Committee on Ocean Science and Resource Management Integration (ICOSRMI). In collaboration with Cascadia Research Collective, the Marine Mammal Institute (MMI), Oregon State University (OSU) is assisting with the integration of photo-identification records and associated genetic samples, to improve understanding of long-term impact of satellite tagging. The resulting database should be suitable for implementation in *geneGIS.*

Title: '*The Shepherd Project'*. This project started as a collaborative software platform for globally coordinated whale shark research, as described in the http://www.ecoceanusa.org/. The success of this platform in managing and supporting the growth of the whale shark catalog led to its selection for the web-based implementation of the SPLASH Photo-ID Catalog (http://www.splashcatalog.org). Through ongoing development of this open-source platform, the *Shepherd Project* provided for the cross application of new functionality to other long-term studies of individually identified marine mammals or marine megafauna.

Title: '*ecoSPLASH*'. This project has provided stable isotope profiles (Carbon and Nitrogen) derived from more than 1,000 skin biopsy samples collected during the SPLASH program. These ecological markers profiles are the linked to individual photo-identification and DNA profiles in the *SPLASH/geneSPLASH* database and have proven informative in descriptions of population structure and the trophic ecology of humpback whales in the North Pacific (Witteveen *et al.* 2009; Witteveen *et al.* 2011a; Witteveen *et al.* 2011b). An initial importation of ^{15}N values shows dramatic regional and individual differences, reflecting difference in the dietary proportions of fish and krill (Figure 7). Inclusion of isotope profiles in the *geneGIS* computational environment provides the potential for more sophisticated analyses of prey specialization and genetic differentiation among feeding regions.

REFERENCES

Calambokidis, J., E.A. Falcone, T.J. Quinn, A.M. Burdin, P.J. Clapham, J.K.B. Ford, C.M. Gabriele, R. LeDuc, D. Mattila, L. Rojas-Bracho, J.M. Straley, B.L. Taylor, J.U. R., D. Weller, B.H. Witteveen, M. Yamaguchi, A. Bendlin, D. Camacho, K. Flynn, A. Havron, J. Huggins and N. Maloney. 2008. SPLASH: Structure of Populations, Levels of Abundance and Status of Humpback Whales in the North Pacific. U.S. Dept of Commerce, Western Administrative Center, Seattle, Washington.

Constantine, R., J. Jackson, D. Steel, C.S. Baker, L. Brooks, D. Burns, P. Clapham, N. Hauser, B. Madon, D. Mattila, M. Oremus, M. Poole, J. Robbins, K. Thompson and C. Garrigue. 2012. Abundance of humpback whales in Oceania using photo-identification and microsatellite genotyping. Marine Ecology Progress Series 453:249-261.

DeSalle, R. and G. Amato. 2004. The expansion of conservation genetics. Nature Reviews Genetics 5:702-712.

Etherington, T.R. 2011. Python based GIS tools for landscape genetics: visualising genetic relatedness and measuring landscape connectivity. Methods in Ecology and Evolution 2:52-55.

Frasier, T.R., P.K. Hamilton, M.W. Brown, L.A. Conger, A.R. Knowlton, M.K. Marx, C.K. Slay, S.D. Kraus and B.N. White. 2007. Patterns of male reproductive success in a highly promiscuous whale species: the endangered North Atlantic right whale. Molecular Ecology 16:5277-5293.

Frasier, T.R., P.K. Hamilton, M.W. Brown, S.D. Kraus and B.N. White. 2009. Sources and Rates of Errors in Methods of Individual Identification for North Atlantic Right Whales. Journal of Mammalogy 90:1246-1255.

Holmberg, J., B. Norman and Z. Arzoumanian. 2008. Robust, comparable population metrics through collaborative photo-monitoring of whale sharks *Rhincodon typus* Ecological Applications 18:222-233.

Lord-Castillo, B.K., B.R. Mate, D.J. Wright and T. Follett. 2009. A Customization of the Arc Marine Data Model to Support Whale Tracking via Satellite Telemetry. Transactions in GIS 13:63–83.

Madon, B., O. Gimenez, B. McArdle, C. Scott Baker and C. Garrigue. 2011. A new method for estimating animal abundance with two sources of data in capture–recapture studies. Methods in Ecology and Evolution:no-no.

Peakall, R. and P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-229.

Rousset, F. 2008. Genepop,'007: a complete re-implementation of the Genepop software for Windows and Linux. Molecular Ecology Resources 8:103-106.

Vandergast, A.G., W.M. Perry, R.V. Lugo and S.A. Hathaway. 2012. Genetic landscapes GIS Toolbox: tools to map patterns of genetic divergence and diversity. Molecular Ecology Resources 11:158-161.

White, G.C. and K.P. Burnham. 1999. Program MARK: Survival estimation from populations of marked animals. Bird Study 46 (Supplement):120-138.

Whitehead, H. 2009. SOCPROG programs: analysing animal social structures. Behavioral Ecology and Sociobiology 63:765-778.

Wieczorek, J., D. Bloom, R. Guralnick, S. Blum, M. Doring, R. Giovanni, T. Robertson and D. Vieglais. 2012. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7:e29715.

Witteveen, B., K. Wynne and J. Roth. 2009. Population structure of North Pacific humpback whales on feeding grounds as shown by stable carbon and nitrogen isotope ratio. Marine Ecology Progress Series 379:299–310.

Witteveen, B.H., J.M. Straley, E. Chenoweth, C.S. Baker, J. Barlow, C. Matkin, C.M. Gabriele, J. Neilson, D. Steel, O. von Ziegesar, A.G. Andrews and A. Hirons. 2011a. Using movements,

genetics and trophic ecology to differentiate inshore from offshore aggregations of humpback whales in the Gulf of Alaska. Endangered Species Research 14:217-225.

Witteveen, B.H., G.A.J. Worthy, K.M. Wynne, A.C. Hirons, A.G. Andrews_III and R.W. Markel. 2011b. Trophic Levels of North Pacific Humpback Whales (*Megaptera novaeangliae*) Through Analysis of Stable Isotopes: Implications on Prey and Resource Quality. Aquatic Mammals 37:101-110.

Wright, D.J., M.J. Blongewicz, P.N. Halpin and J. Breman. 2007. Arc Marine: GIS for a Blue Planet. ESRI Press..

CONFERENCE PRESENTATIONS

Dick, D., D. Wright, J. Calambokidis, J. Holmberg, E. Falcone, T. Follett, D. Steel, B. Slikas, and C. S. Baker. 2012. *geneGIS*: computational tools for spatial analyses of DNA profiles associated and photo-identification records of whales and dolphins. Poster Presentation, The Wildlife Society 19th Annual Conference, Portland, OR, 13-18 October 2012.

Holmberg, J. 2012. *The Shepherd Project:* A software framework for mark-recapture and molecular ecology. Oral Presentation, The Wildlife Society 19th Annual Conference, Portland, OR, 13-18 October 2012.